10 Database and File Concepts

- A database is a program used to store data in structured way
- This includes the data that is stored and the links between the data items
- All databases store data using a system of files, records and fields
- There are two types of database: a flat-file database and relational database

10.1 Flat-file and relational databases

- A flat-file database stores its data in one table, which is organized using rows and columns.
- Each column in the table contains a field which has been given a field name and each cell in that column has the same, predefined data type

- A relational database stores the data in more than one linked table, within the file
- It is designed so that the same data isn't stored many times
- Each table with in a relational database will have a key field.
- Most tables will have a primary key field which holds unique data and is the field used to identify that record
- Some tables will have one or more foreign key fields
- A foreign key in one table will point to a primary key in another table

Why One Might Be Preferred in Certain Situations:

Advantages of Relational Database

- A flat file database can contain many fields, often with duplicate data leading to wasting storage
- If the flat files contain some of the same fields, when data in one is changed the other has to be manually modified, wasting time
- Adding records to a relational database is easier as less data has to be entered saving time
- In a relational database certain tables can be made confidential so that when a person logs on with their username and password, the system can then limit access only to those tables whose records they are authorized to view, e.g. a receptionist would be

able to a view a workers department and be able to locate them but wouldn't be able to see their salary
- Searching a database is quicker to locate records
- It is easier for users to produce reports
- Integrity constraints ensure that table relationships remain valid

Disadvantages of Relational Database

- Designing a relational database will take a lot of planning whereas the fat files already exist
- Creating a relational database requires technical expertise which will have to be paid for
- Some relational databases have limits on field lengths and this can lead to data loss
- If the number of tables increases, setting up the relationship is much easier

10.2 Relationship types

Relationships are used to link tables together.

One-to-One Relationship (1-1)

- This relationship is where each record in one table relates to only one record in another table.
- For this relationship to exist, the linked fields in both data tables must contain the same unique values

One-To-Many Relationship (1- ∞)

- This relationship is where each record in one table can relate to many records in another table
- The primary key in one table relates to the foreign key field in the second table

Many-to-Many Relationship

- A many-to-many relationship has to be broken down into several one-to-many relationships using a third table, a "join table" where each record in the "join table"

would have the foreign key fields linked to the primary keys of the tables it is joining together

10.3 The function of key fields

Key fields are used to create the relationships between different tables in the database. All relationships between tables use key fields. There are three types of them:

Primary Key

- A primary key is a field in a table which is unique and enables you to identify every record in that table
- It must not contain null values and must contain unique value for each row of data
- A table in a relational database must always have one and only one primary key
- It optimizes the storage of table records within the database and helps to provide quicker research

Compound Key

- A compound key is a primary key that combines more than one foreign key to make a unique value/primary key
- It consists of more than one field

Foreign Key

- A foreign key is a primary key field in one table that is linked to the primary key in another table
- It is used to form a relationship between the two tables
- Referential integrity is usually implemented through the use of foreign keys
- There can be many foreign keys in a table and can contain duplicate values

10.4 Referential Integrity

- Referential integrity forces table relationships to be consistent and avoids redundant data

- This means that the data in the foreign key field must match the data in the primary key field
- It prevents records being added to a related table if there is no associated record in the primary table
- It prevents a user from accidentally changing or deleting data in one table, without the same action happening to the related data in another table

10.5 Normalisation

- Normalisation is a method of organizing data tables which involves breaking down a flat-file
- It attempts to avoid data loss
- It attempts to avoid data redundancy (duplication)
- It attempts to improved data base efficiency
- This technique is a multi-step process, where each step has a rule that improves the efficiency of the database
- These rules are called Normal Forms such as 1NF, 2NF and 3NF.

Un-normalised Form (0NF or UNF)

- If a database is not normalized, it is called an un-normalised database often called 0NF or UNF,
- This is usually a flat-file database that contains duplicated data and complex data structures (non-atomic data) stored within a single field.

First Normal Form (1NF)

The rules for a database normalized to 1NF are:

- All data is stored in a database table
- A unique key must exist in each table
- Only atomic data (which can't be broke down any further) is stored
- Each field has a unique name
- Each record is unique (removes duplicated records)
- There are no repeating groups of columns

Second Normal Form (2NF)

The rules for a database normalized to 2NF are:

- Table must be in 1NF
- Related data needs to be separated into different tables
- Each table should have a primary key
- The fields in each table should be dependent on the primary key *(any non-key attributes that only depend on part of the table keys are placed in a new table)*

Third Normal Form (3NF)

The rules for a database normalized to 3NF are:

- The table should be in 2NF
- The database shouldn't have any non-key dependencies *(any non-key attributes that are dependent on other non-key attributes should be removed and placed to another table)*

Advantages of Normalisation

- Any change to one record which is needed an instantly be made to any related records
- The database doesn't have redundant data making the file size smaller so less money needs to be spent on storage
- There is no data duplication so there are fewer errors in the data
- There is no data duplication so there is less chance of storing incorrect copies of the data
- There is no data duplication making the file size smaller so less money needs to be spent on storage
- Modifying a table is easier as there is less data to modify

Disadvantages of Normalisation

- A large number of tables requires ore relationships to be designed taking more time
- Making data atomic may not always be the best solution such as date of birth can be separated into day, month and year but this may serve no purpose
- Data may be stored as odes rather than meaningful data making it difficult for humans to interpret it
- With more tables, setting up queries can become more difficult, the complex the query the longer it can take to run

- You can end up with more tables than an unnormalized database making it difficult to keep track of data
- May require greater expertise which may need to come from outside and be paid for, creating extra expense

10.6 Data dictionary

- A data dictionary is a file containing description of and information about the structure of data in a database
- It is often designed before a database is created as an alternative to an entity relationship diagram

Components of a Data Dictionary

The data dictionary should always contain the:

- Table names- unique name for each table in a database
- Field names- to identify each field
- Data types- e.g. text, numeric, Boolean
- Field size- the number of characters in each field
- Keys- where the primary and foreign key fields are identified
- Other metadata which could include input masks, validation rules or default values

Table name: Jobs				
Field name	**Data type**	**Field size**	**Key field**	**Metadata**
JobCode	Alphanumeric/Text	3	PK	
Description	Alphanumeric/Text	28		

10.7 Query Selection

Static Parameter Query

- Static parameter query is where the parameter that the query uses has been set as a fixed value within the query
- In a static query, every time the query is run it will search for the same data
- It can only be changed by recreating the query or editing its structure

- It is used when criteria need embedding into query, so that the same query is run repeatedly even if the data in the table changes

Dynamic Query

- Dynamic parameter query is a query that prompts the user for input to search for different values each time it is run
- The input is then used by the query as the value in an expression or criteria
- Every time the query is run a dialogue box would appear asking the user to type in the value
- It is used when different values need to be searched each time or when the data has to be entered by the user

Compare and contrast static parameter query with dynamic parameter query [5]

- Both search for the data that matches the search criteria
- Both return records containing the data
- Both can be used with complex queries
- A dynamic query can be used to search for different values each time it is run
- In a static query every time that the query is run it will search for the same values
- Dynamic query needs a value to be entered every time
- Static query has the criteria hard coded hence doesn't need to be added each time
- Dynamic query would save the time of designing the query every time a different criteria was used
- Dynamic query requires more technical knowledge of the user and hence is more complicated to create

Simple Query

- Simple query is one that searches using a single criterion only
- Simple query is used only on one field

Complex Query

- A complex query is one that searches using more than one criterion
- It often uses Boolean operands

- It is made up of AND, OR or NOT operands or a mixture of these
- It is used on more than one field

Nested Query

- A nested query is a query within another query, often referred to as a subquery
- Nested queries help you to use the result of one query as an input parameter of another
- Innermost subquery is executed first, then next level, until the main query is reached
- It is used when data from one query needs to be used for a different type of query such as producing a query to summarize data based upon the results from a simple or complex query

Summary Query

- Summary queries are used to summarize the contents of a table
- Also called Group-By queries/aggregate queries and use aggregate functions
- Uses functions such SUM, AVERAGE, MIN, MAX, COUNT
- A crosstab query is a type of summary query which displays results in a similar format to a pivot table in a spreadsheet

10.8 File and data management

13

Program files

- Program files are saved with .exe file extension, they are often opened by clicking on a program icon

Proprietary file formats

- Proprietary file formats are formats belonging to a company, organization or individual and are often specific to a program created and copyrighted by that company.

- These files contain your data but are stored in the formats of the particular package and contain more information than just the contents that you can see when the document is opened.
- The exact details of the data stored, the encoding and its structure within the file are often kept secret and patented by the company or organization that created the software.
- Proprietary software allows only people with licenses to use it
- A proprietary file format has been created by a company using a particular encoding scheme
- Designed by the company such that the decoding of this stored data can easily be done with software that the company itself developed
- The specification of the encoding is usually kept a secret through licenses such as that the only the company may use it

Open-source file formats

- These are formats used for storing data that are published for anyone to use.
- They can be opened by both proprietary and open-source software.
- Anyone can use it as no monetary cost is required
- The details of the way the data is stored are available for all users and software developers rather than just the organization which created the structure.
- The code is open to inspection and sometimes to amendment without breach of copyright legislation.
- They are called free file formats if they aren't covered by any copyrights
- Users can transfer files from a work computer to a home computer using open-source software more easily, as computers being used may not have compatible formats
- They can be opened by most types of software
- There is published specification for storing digital data, usually maintained by a standards organization

Generic file formats

- Generic file formats allow you to save files so that they can be opened on any platform, for example files created on a PC can be read/imported on a mobile phone, etc and vice versa.
- If a software package doesn't recognize a particular file format it will be unable to load it

- Generic file formats are .txt that can be loaded by most word processing software, .csv, etc
- However, the files may not contain all the formatting that can be saved in a package-specific format and it is not always possible to open proprietary file formats on other platforms. *Do q53& w23/12/q8*

Common generic text files include:

- **Comma separated values:** These files have a **.csv** file extension, it takes data in the form of tables and saves it in text format, separating data with commas
- **Text:** These files have a **.txt** file extension, a text file isn't formatted and can be opened in any word processor software
- **Rich text format:** These files have an .rtf file extension, this is a text file type that saves some of the formatting within the text

Common generic image files include:

- **Graphics interchange format (GIF):** These files have a **.gif** file extension that stores still or moving images, it is an efficient method of storing images using a smaller file size where there are large areas of solid color, mostly used in webpages
- **Joint photographic expert group (JPEG):** These files have a **.jpg** or **.jpeg** file extension which stores still images but not moving ones, it is an efficient method of storing images using a smaller file size and is used in web pages
- **Portable document format (PDF):** These files have a **.pdf** extension. This is a document which has been converted into an image format, it allows documents to be seen as image so they can be read on most computers, the pages look kusti ke they would if they were printed but can contain clickable links and buttons, form fields, video and audio. In pdf format, you can protect a document to stop others from editing it
- **Portable network graphics (PNG):** these files have a **.png** file extension, it is a format that compresses graphics files (image) without any loss of image quality, it is the most used lossless image compression format on the internet
- **Moving pictures experts group layer-4 (MPEG-4):** these files have an **.mp4** file extension, its not a single file format, it's a multimedia container which is used for storing video files, still images, audio files, subtitles and so on. It is used to transfer video files on the internet

Common generic audio files include:

- **Moving pictures experts group layer-3 (MPEG-3):** These files have an **.mp3** file extension, it's a compressed file format used for storing audio files, it cant store still or

moving images. File sizes are relatively small but have high quality, which makes it suitable for use on the internet.

Common generic files used for website authoring:

- **Cascading style sheet:** These files have a **.css** file extension, this is a stylesheet which is saved in a cascading stylesheet format and is attached to one or more webpages (usually HTML) o define the age's color, scheme, fonts and so on.
- **Hypertext markup language (HTML):** These files have an **.html** or **.htm** file extension. This is a text-based language used to create makeup that a web browser will be able to interpret to display information on a webpage

Common generic compressed files include:

- **Roshal archive (RAR):** these files have **.rar** file extension, which can hold almost any file type in a compressed format. Used to reduce the number of bytes needed to save a file, either to save storage space or to reduce transmission time.
- **Zip:** These files have a **.zip** file extension which can hold almost any file type in a compressed format, it is used to reduce the number of bytes needed to save a file, either to save storage space or to reduce transmission time

10.9 Sequential Access

- Records are accessed in the order they were entered
- Each record is read one by one until a match is found
- The only way to add new records to a sequential file is to store the at end of file
- A record can only be replaced if the new record is exactly the same length as the original
- Records can only be updated if the data item used to replace the existing data is exactly the same length
- The processing of records is also slower in a sequential file, as it has to read all records in sequence before the one that is required.
- Use of sequential file is only recommended for applications where most or all the records have to be processed at one time

10.10 Indexed Sequential Access

- Indexed sequential access is a mixture of sequential and direct/random access

- Indexed sequential files are stored in order
- They are stored on disk to allow direct access unlike sequential access files which are stored on tape
- Each record consists of fixed length fields
- It allows records to be accessed either sequentially in the order they were entered or randomly using an index
- Each index defines a different ordering of the records
- A database may have several indexes, based on the information required
- A key is specified in each index
- it is a method of indexing data for fast retrieval.
- A set of hash tables known a s indexes contain “pointers” into the records
- Individual records can be retrieved without having to search the entire file
- Indexes can be searched quickly, thereby allowing the database to access only the records it needs
- Searches use an index which will narrow down the records to be searched
- Then that section of the file is searched sequentially to find the record required
- The use of ordering enables fast access as a table of indexes is used to allow the search to jump to a particular place on the disk rather than going through all the other records to get to the right point.

10.11 Direct Data/Random Access

- Random access is the quickest form of access
- The whereabouts of records in file doesn’t matter, it will take same amount of time to access any particular record
- Each record is fixed length and has a key
- The computer looks up the key and goes to the appropriate place on the disk to access it

10.12 Hierarchical Database Management System 42

- It has a tree like structure, where data is stored as records connected to one another through links
- It links a number of records to one primary key
- It uses several one-to-many relationships
- It isn’t a versatile system
- It is limited by using only one type of relationship so confined to specific uses

- This allows fast access to data, because large amounts of data are bypassed as you go down the level of its structure

10.13 Features of a Management Information System (MIS)

- MIS is a computer-based system that provides managers with the tools to organize, evaluate and efficiently manage departments within an organization
- It provides past, present and prediction information
- It includes software that helps in decision making
- It includes (many data resources such as) databases
- It includes hardware resources of a system
- It includes decision support systems, people management and project management application

10.14 How a MIS can be used by organizations

- An MIS can help a company to run more efficiently because they provide details about the past and present performance of the company and give prediction for its performance in future.
- It helps managers in decision making
- MIS manager typically analyses business problems
- MIS manager designs and maintains computer applications to solve the organizations problems
- Its helps with project management
- Managers use management information systems to gather and analyses information about various aspects of the organization such as personnel, sales, inventory and production
- MIS are used to create report (on as aspects such as sales, revenue and product)
- These reports are provided at regular intervals to managers at all levels to help them evaluate their company's performance
- By comparing daily, weekly or monthly reports to previous reports, managers are able to spot trends such as revenue growth or reduction
- By creating charts, can see trends such as revenue growth or reduction